UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK
---x
UNITED STATES OF AMERICA 14-CR-414 (BMC)

-against-

RASHAWN JERMAINE SMALLS
---x

DECLARATION OF DR. RANAJIT CHAKRABORTY

I, DR. RANAJIT CHAKRABORTY, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I am a professor at the Department of Molecular and Medical Genetics and the Director of the Center for Computational Genomics at the Institute of Applied Genetics at the University of North Texas Health Science at Fort Worth, Texas.
2. Among the courses I teach are courses in statistical genetics, forensic data analysis, molecular biology and genetics, and testifying as an expert witness. I also conduct research on subjects including current issues of DNA forensics, using new technologies involving bioinformatics and data modeling.
3. My curriculum vitae, attached to this declaration as Exhibit A, lays out my credentials, including my educational background in the field of statistics and genetics; my numerous publications relating to forensic DNA testing methodology and population genetics; and the scientific journals on whose editorial boards I serve.
4. One significant forensic application of my research includes my assistance in developing the 13 core short tandem repeats ("STRs") genetic markers used by the FBI and in labs throughout the world. In their 1996 report on DNA Forensics, the US National Research Council cited my research more than two dozen times as foundations for the recommendations made in their report.

5. I served on the New York State DNA Subcommittee from its inception in 1995 until 2011. The DNA Subcommittee grants accreditation for all forensic laboratories in New York State. The Subcommittee also advises the Commission on Forensic Science on any matter related to the implementation of scientific controls and quality assurance procedures for the performance of forensic DNA analysis.

6. As part of my work on the DNA Subcommittee, I reviewed the materials that the Office of the Chief Medical Examiner ("OCME") of the City of New York supplied to the subcommittee in 2010 regarding the Forensic Statistical Tool ("FST"). These materials contained summaries of their validation studies done until 2010, examples of data sets of different components of the validation studies, and PowerPoint presentations from the OCME. OCME did not provide the DNA Subcommittee members with the source code for the FST, or with live demonstrations of the actual FST program.

7. In 2010, as a member of the DNA Subcommittee, I approved the FST for use in forensic casework for the purposes for which it was validated at that time, but I also expressed some concerns about the methodology underlying the FST and recommended that additional work be done. Since voting to approve the FST, I have developed serious concerns about the use of the FST based on OCME's subsequent publications, and additional research and developments in the field, particularly in relation to application of the FST in complex cases.

8. I have been qualified as an expert in over 200 cases in the areas of DNA typing methodology and interpretation, including the area of population genetics. In over 95% of those cases, I have testified on behalf of the prosecution. In about a dozen of those cases, I have testified on behalf of the defense. Prior to my involvement in this case, I

testified in two other cases involving use of the FST. On both occasions, I testified for the defense.

9. In December 2013, I testified at a *Frye* hearing as an expert in forensic DNA analysis, population genetics, and statistical genetics in the case of *People v. Peaks & Collins*, Ind. Nos. 7589/2010 & 8077/2010, Supreme Court of New York, Kings County, about the FST. In that case, the Court determined that the FST has not gained general acceptance in the relevant scientific community. It is my understanding that within the last month the Honorable Judge re-asserted his ruling in spite of the prosecutor's appeal to re-open the admissibility hearing based on recent developments in the field.

10. In connection with this declaration, I have reviewed the OCME's reports and work papers generated from DNA samples taken from: (1) Rashawn Jermaine Smalls, (2) an alleged witness, and (3) the gun recovered from the scene. These reports are attached as Exhibits C-E to the Colson Affirmation. I have also reviewed OCME's validation studies and its protocol for use of the FST. The protocol is attached as Exhibit J and the studies are attached as Exhibits K-L to the Colson Affirmation.

11. Upon reviewing OCME's materials, and from my extensive experience with the FST, including in my role as a member of the DNA Subcommittee that reviewed the FST, I have concluded that the FST, as used in this case, is unreliable and not fit for use in real-world criminal cases such as the present one. Hence, the FST results in this case are unreliable and prejudicial to Mr. Smalls.

DNA Profiling

12. DNA (deoxyribonucleic acid) is the hereditary material found inside all of us that makes us human. It is frequently referred to as our "genetic blueprint." DNA is located in every

nucleated cell in the body. No two individuals, with the exception of identical twins, will have the same totality of genetic codes.

13. A chromosome is the tightly packaged structure of DNA, present within nucleus of every cell having a nucleus. Each such individual human cell contains 22 pairs of chromosomes; one chromosome from each pair is inherited from the individual's father and the other comes from the individual's mother. There is also a 23rd pair, which is called the sex chromosome, X and Y.

14. Most modern forensic analyses look at Short Tandem Repeats ("STRs"), which are small segments of DNA that are repeated in tandem. STRs are distributed among each person's DNA at specific locations ("loci").

15. An allele is a number of repeats of the STR motif which describes the size of the DNA fragment, or the number of STRs, at a location ("locus"). Each individual has two alleles at each locus—one inherited from the mother and one from the father. If the two alleles at a locus are different, they are known as "heterozygous." If they are the same, then they are "homozygous."

16. Generally, 13 core loci are examined, with some laboratories testing two additional loci. One number is reported for each allele at a locus. For example, if one allele has 13 repeats of the STR motif and the other has 18 repeats, the result is reported at 13, 18 for the particular locus. An individual DNA profile is a string of numbers representing the two alleles at each of the loci examined. Although many individuals will share a particular allele at a particular locus, the chance that two people will have the same set of alleles at multiple loci decreases as the number of loci examined increases. Thus, an

analyst can use the observed allele combinations at multiple loci to distinguish one individual from another.

17. Basic forensic DNA analysis involves several steps. In the first step, "extraction," DNA is extracted from the evidence, the gun in this case. In the second step, "quantification," the analyst measures the amount of DNA present in the sample. The third step is "amplification," in which a process called polymerase chain reaction ("PCR") is applied to the DNA sample. PCR makes millions of copies of a particular segment of DNA so that it can be detected and analyzed. In the fourth step, a process known as "electrophoresis" separates the STR fragments by size. That data then becomes the input for a software program—in this case Genemapper—that converts the data to graphs which can be formed and presented in a number of ways determined by the operator. The labeling and graphs produced by the Genemapper program are called electropherograms ("epgs").

18. Once an electropherogram is generated, the analyst reviews it, draws conclusions about the DNA sample, and creates a DNA profile. In the final step, the analyst compares the profile created with the suspect's DNA profile.

Forensic DNA Analysis as Applied to DNA Mixtures

19. A DNA mixture is defined as a DNA profile that contains DNA from multiple individuals or contributors, for example, the gun swab in this case.

20. DNA extracted from an evidence sample containing the DNA of multiple contributors cannot, by itself, reveal the DNA profiles of each contributor.

21. While the step of quantification provides an estimate as to the total amount of DNA in a mixture, the specific quantity of DNA from each contributor remains unknown, irrespective of the accuracy of the quantification process.

22. PCR amplification is a sequential process, namely, at each cycle of amplification, DNA molecules are doubled in number. Preferential amplification of DNA from any individual contributor to a DNA mixture may create imbalance in the number of amplified DNA molecules from different contributors. This is particularly important for contributors whose contribution to the mixture is substantially lower than others (resulting in the event of allele drop-out, discussed in the next section below).

23. The mixture ratio is the proportion of DNA contributed by each contributor to the DNA mixture. The mixture ratio has a critical impact on the interpretation of data during electrophoresis. At this time, the amplified DNA molecules of the sample interact with the reagents of commercial kits used in the capillary electrophoresis process to provide signatures of alleles present in the amplified DNA molecule. The amount of these reagents is finite. Hence, there is a competition from the different alleles at a locus to react with the reagents. Alleles that are present in a fewer number of molecules of the amplified product will obviously leave a smaller signature of their presence in the DNA mixture. This leads to an imbalance of peak heights in the electropherograms (epgs). Thus, at all loci, the signatures of alleles present in an individual who contributed less DNA to the mixture are expected to be weaker than those of the alleles of major contributors. Hence, peak heights by themselves, also become relevant in interpreting DNA profiles, particularly in DNA mixture analysis.

Drop-In and Drop-Out

24. The process of PCR can result in random errors or "stochastic effects." Two of the most common errors are allele drop-out and allele drop-in. Allele drop-out occurs when alleles from the principle DNA donors fail to appear in the profile. Allele drop-in occurs when alleles not originating from the principle DNA donors to a mixture show up in a DNA profile. The concepts of drop-out and drop-in are relevant for single source DNA as well as DNA mixtures and can be even more complex when DNA is degraded.

25. Dirt, bacteria, and sunlight can all cause DNA degradation. Touched DNA is also frequently degraded. DNA varies by length, and longer pieces of DNA will break down faster than shorter pieces. That means, when DNA is degraded, some longer pieces of DNA may not be detected.

26. One additional reason for drop-out involves the relative proportions of a DNA mixture, as mentioned above. If the mixture is not even (as reflected in mixture ratios), and if one person is represented in greater amounts, more pieces of one contributor's DNA might be grabbed during sampling as well as during electrophoresis, and therefore some DNA components (alleles) would not be detected.

The Forensic Statistical Tool

27. The FST is a proprietary software program developed by the New York City Office of the Chief Medical Examiner ("OCME"). The program examines the alleles found in DNA mixtures that cannot be "deconvoluted," or separated into individual profiles. Then it determines a statistical weight, or likelihood ratio ("LR") that provides the relative support of the scenario that the suspect's DNA is included in the mixture. The numerator

of the LR includes the data favoring the prosecution scenario, while the denominator includes the data favoring the defense scenario.

28. There are various forensic software programs that calculate LR's, and each has it benefits and drawbacks. Research on this topic is still in its infancy, and in my opinion, none of these programs is ready for use in court at the present time.

29. The FST is unique among LR programs in that it uses pre-determined drop-in and drop-out rates based on DNA quantity. There are a number of problems with the way that the OCME arrived at the drop-out rates it uses.

30. As an initial matter, OCME has still not released the source code for the FST program or made the program available for public use. Without knowing the source-code, it is virtually impossible to test the reliability of the LR calculations performed in this case. OMCE's publications to date are insufficient to assess reliability.

31. In addition, there is a 30% range of error in measuring DNA quantity. To my knowledge, OCME has not published any studies explaining how the error rate may affect the FST. In my opinion, however, a 30% error rate is unacceptable for use in estimating drop-out rates because the FST provides only *one* statistical value. OCME's failure to test or identify how the 30% error rate impacts the LR generated by the FST undermines the program's reliability. Furthermore, with the quantity of DNA from each contributor being unknown from the DNA quantity of the whole mixture, the use of a single (point) estimate of the DNA quantity in the whole mixture has little, if any, relevance on the drop-out rate of specific alleles contributed by a specific (presumed) contributor. This issue is even more critical, since the FST software does not estimate or invoke the mixture ratios by using case-specific data.

32. Validation studies are the process by which laboratories demonstrate that their instruments are robust and reliable. The DNA samples examined in OCME's validation studies were extracted from "pristine buccal swabs," and OCME assumed a minimal number of possible mixture ratios. By contrast, DNA extracted from evidence in the field is most often degraded, and the mixture ratios are unlikely to conform to those derived from the validation studies. The validation studies also ignored the possible relatedness of the contributors to each other or the suspect, and the effect of relatedness that may have on allele composition of the DNA profile of a mixture sample. These variables affect the drop-in and drop-out rates on a case-specific basis and thus affect the range of reasonable LRs. *Every* criminal case is influenced by at least one of these variables.

33. Despite OCME's own admissions of weaknesses in the FST program, to date, OCME has not published any additional studies documenting how case-specific variables may affect the LR. Nor has OCME made any documented adjustments to the FST program.

34. During my review of OCME's FST analysis of DNA mixtures, I expressed concerns about how the allele drop-out rates for alleles of any single presumed contributor varied from one locus to another, which is technically called the "independence of drop-out rates" of alleles originating from the same contributor.

35. Specifically, the OCME noted in its conclusions that: "Drop-out rates appear to be independent across loci. That is, drop-out or lack of drop-out at each locus <u>is not consistently associated</u> with an increased or decreased probability of drop-out at other loci."). *See* Likelihood Ratio Statistics For Analysis of Single Source, Mixed and Degraded Evidence Samples, Volume 22: Determination of Independence of Drop-Out Among Loci, Summary at 1 (emphasis added). This analysis is attached as Exhibit B.

This indicates that there were some associations, and that the conclusion that drop-out rates appear to be independent across loci was incorrect.

36. In addition, the data used by OCME to determine independence across loci was limited and does not account for all of the real-life factors that could influence the dependence or independence of drop-out rates across alleles. These include the varying number of contributors, allele sharing between them, varying mixture ratios, and uneven degradation of their DNA. Nor does OCME factor its failure to consider these variables into its calculation of drop-out rates.

37. Also, OCME's testing of drop-out across loci was too simplistic. OCME chose one locus, and asked whether drop-out rates there were consistent with drop-out at other loci under various conditions. Instead, it should have considered whether drop-out rates of all 15 loci were simultaneously independent of each other. In short, OCME's validation studies are inadequate and do not include important analyses that should have been performed.

38. I made suggestions that further work be done on the dependence of drop-out rates across loci, but to my knowledge, the OCME has not followed my suggestions to do this additional work. In my opinion, this failure to do further work on the dependence of drop-out rates renders further support to claim that the FST as applied in this case is unreliable.

39. After the DNA Subcommittee approved the FST, I also learned that, while mixture samples that were created for testing the validity of the FST protocol included contributors from difference races and ethnicities, that information was not used to test the independence of drop-out rates. This is important because allele frequency is an

important possible predictor of drop-out rate, and allele frequencies vary by race and ethnicity. If race and ethnicity are not properly accounted for, the results will not be reliable.

40. Data reflecting the racial characteristics of the contributors were not explicitly presented to the DNA Subcommittee, nor was the data preserved.

41. Empirical multi-locus data (i.e., multiple locus genotype data on actual sampled individuals) used in some of OCME's validation studies did not include two additional loci—loci D2S1338 and D19S433—since these were not included in the commercial kit used for DNA profiling at OCME's sampled individuals. Based on published allele frequency data, OCME simulated genotypes for these two additional loci and included them in the multi-locus DNA profile data in their validation studies. OCME did not reveal this approach during its presentations to the DNA Subcommittee. Moreover, it is statistically unacceptable because it produces pseudo-independence of genotypes of some loci in their data. If this fact was known during my participation on the DNA Subcommittee, my decision could have been drastically affected.

42. The OCME has also introduced further uncertainties in its calculation of LRs using the FST by underestimating the drop-out rates. OCME assumes that lower drop-out rates make the LR lower, favoring the defense. The OCME characterizes this as a "conservative" approach.

43. However, OCME's validation studies show that underestimating drop-out rates only leads to a lower LR in approximately 50% of the cases. Whatever additional studies OCME may have done to justify a differing conclusion are unrecorded and unreported.

No mention of the studies was made to the DNA Subcommittee or reported in scholarly articles.

44. The OCME also underestimates the number of contributors to the mixture under the theory that this is a "conservative" approach. This premise is flawed, and the impact of unreliability of LR calculations with inappropriate assumptions on the number of contributors has been extensively discussed in the literature. *See, e.g.,* H. Haned, CCG Benchop, P.D. Gill, T. Sijen, *Complex DNA mixture analysis in a forensic context: Evaluating the probabtive value using a likelihood ratio model*, Forensic Science Int'l: Genetics (2015). This article is attached as Exhibit R to the Colson Affirmation.

FST As Applied To This Case

45. I have reviewed the OCME's reports in this case, the final one of which compares the mixture profile of the swabs from the "entire grip area" and "slide slip grooves" of the gun with the known profile of Jermaine Smalls.

46. OCME performed two electropherogram runs on the DNA mixture from the grip area. The two runs produced different results at 13 of the 15 loci examined because of allele drop-in and drop-out. For example, at locus D19S433, Run One identified six alleles, and Run Two identified seven alleles. A total of nine alleles were observed from both runs, indicating that at least four people likely contributed to that locus. In addition, no alleles were present on either run at locus D2S1338, indicating full drop-out at that locus.

47. Because the two runs showed different results, OCME compared all of the alleles from both runs with Mr. Smalls' DNA profile.

48. Using the FST, OCME then concluded that:

(a) Jermaine Smalls cannot be ruled out as a contributor;

(b) the swab from "entire grip area" is approximately 4,190 times more probable if the sample originated from Jermaine Smalls and two unknown, unrelated persons than if it originated from three unknown, unrelated persons. These results were also interpreted as "very strong support" that Jermaine Smalls and two unknown, unrelated persons contributed to this mixture; and

(c) "[b]ased on the number of DNA alleles seen in the DNA profile of Jermaine Smalls that are not seen in the mixture of DNA found in the [slide slip grooves], it cannot be determined whether they can or cannot be excluded as a contributor to the mixture."

49. I have concerns about the reliability of these results. First, according to OCME's own validation studies, the mixture extracted from the grip area included the DNA of at least four people. The greatest number of alleles a single person can have across 15 loci is 30. Most individuals have fewer than 30 alleles because they may be homozygous at certain locations. According to OCME's validation studies, the maximum number of alleles OCME expects to see in a three-person mixture is 64. Where there are more than 67 alleles in a mixture, OCME considers it a four-person mixture. The forensic sample from the grip area contained 69 alleles. Thus, according to OCME's own guidelines, it should have viewed the evidence as a four-person mixture. Had it done so, the LR undoubtedly would have differed. To my knowledge, however, OCME has not performed validation studies on four-person mixtures. Nor has the FST been specifically approved for use on four-person samples. Thus, the reliability of the FST as applied to four-person mixtures is unknown.

50. Second, the DNA mixture extracted from the grip area was degraded. An examination of the electropherogram of the grip area sample shows a classic "ski slope" effect. The "ski slope" effect is the decrease in size of the peak heights from left to right across the allelic positions. In the grip area electropherogram, for example, the peak heights are nearly 70% higher on the left than the right. The "ski slope" effect is a well-accepted indicator of degradation among the scientific community. To my knowledge OCME has not performed validation studies on three- or four-person degraded mixtures. Thus, the reliability of the FST as applied to degraded mixtures is unknown.

51. Third, the FST protocol directs that the program should be used whenever "[t]he DNA profiles of the major and the minor contributors cannot be determined; however, the sample is informative and suitable for comparison." The sample from the slide slip grooves was found suitable for comparison. Thus, according to its own protocol, OCME should have used the FST on that sample. Having failed to do so, there is no basis for stating it "cannot be determined" whether or not Mr. Smalls contributed to that mixture.

52. Fourth, the LR computation does not explicitly delineate how allele drop-out or drop-in has been modeled in these computations. Nor it is stated how allele drop-out at different loci were treated as possible dependent events. Again, without the source code of FST unavailable, and without any information on what parameters were used for this specific application of FST in this case, any claim of reliability of the LR calculation of this case cannot be verified, rendering this report unacceptable.

53. In interpreting the results generated by the FST, the OCME has also developed a table for expressing the strength of support for one hypothesis over another. The table is set forth below.

Reported Value	**Qualitative Interpretation**
Less than 0.001	Very strong support for Defense Hypothesis over Prosecution Hypothesis
0.01 to 0.001	Strong support for Defense Hypothesis over Prosecution Hypothesis
0.01 to 0.1	Moderate support for Defense Hypothesis over Prosecution Hypothesis
0.1 to 1	Limited support for Defense Hypothesis over Prosecution Hypothesis
1	No conclusions
1 to 10	Limited Support for Prosecution Hypothesis over Defense Hypothesis
10 to 100	Moderate Support for Prosecution Hypothesis over Defense Hypothesis
100 to 1000	Strong Support for Prosecution Hypothesis over Defense Hypothesis
Greater than 1000	Very Strong Support for Prosecution Hypothesis over Defense Hypothesis

54. Uncertainties of the LR estimate are not explored in these qualitative interpretations, in the absence of which the use of verbal predictions can be termed at best as subjective stipulations. For example, if the confidence of the LR estimate is 10-fold lower (because of sampling errors of allele frequency estimates, and other uncertainties mentioned

earlier), the evidence may not be appropriately characterized. Considering the uncertainties of the LR estimate, the conclusion that there is "very strong support" that Jermaine Smalls contributed to the mixture is entirely subjective.

55. In view of these factors, the scientific basis of this DNA evidence is very weak, and the failure to express the level of uncertainties in the computations is likely to generate a false support of this weak DNA evidence.

56. I declare under the penalty of perjury that the foregoing is true and accurate to the best of my knowledge, information and belief.

Dr. Ranajit Chakraborty

Dated: January 22, 2015
Fort Worth, Texas